

Predicted highly expressed and putative alien genes of *Deinococcus radiodurans* and implications for resistance to ionizing radiation damage

Samuel Karlin* and Jan Mrázek

Department of Mathematics, Stanford University, Stanford, CA 94305-2125

Contributed by Samuel Karlin, February 15, 2001

Predicted highly expressed (PHX) and putative alien genes determined by codon usages are characterized in the genome of *Deinococcus radiodurans* (strain R1). *Deinococcus radiodurans* (DEIRA) can survive very high doses of ionizing radiation that are lethal to virtually all other organisms. It has been argued that DEIRA is endowed with enhanced repair systems that provide protection and stability. However, predicted expression levels of DNA repair proteins with the exception of RecA tend to be low and do not distinguish DEIRA from other prokaryotes. In this paper, the capability of DEIRA to resist extreme doses of ionizing and UV radiation is attributed to an unusually high number of PHX chaperone/degradation, protease, and detoxification genes. Explicitly, compared with all current complete prokaryotic genomes, DEIRA contains the greatest number of PHX detoxification and protease proteins. Other sources of environmental protection against severe conditions of UV radiation, desiccation, and thermal effects for DEIRA are the several 5-layer (surface structure) PHX proteins. The top PHX gene of DEIRA is the multifunctional tricarboxylic acid (TCA) gene aconitase, which, apart from its role in respiration, also alerts the cell to oxidative damage.

The genome sequence of *Deinococcus radiodurans* [DEIRA (SWISSPROT nomenclature)] strain R1 became available in 1999 (1). DEIRA has a major Chromosome I (2.65 Mb), a megaplasmid-1 or possible "Chromosome II" (412 kb), a megaplasmid-2 (177.5 kb), and a usual size plasmid-3 (45.7 kb). DEIRA is extremely resistant to various agents and conditions that damage DNA, mostly those caused by ionizing and UV radiation. The mechanisms of this resistance are unclear. Our objectives in this paper are to report on predicted highly expressed genes (PHX) and putative alien (PA) genes (see below for formal definitions) and their implications how these genes may help guide research into understanding DEIRA survival through periods of extended ionizing damage and desiccation (2–6). The complete lists of PHX and PA genes in DEIRA are presented in our ftp site, <ftp://gnomic.stanford.edu/pub/highlyexpressed>. For recent reviews on the radiobiology of *D. radiodurans*, see refs. 5 and 7.

Predicted Highly Expressed Genes

Our approach in predicting gene expression levels relates to codon usage differences among gene classes. Genes that deviate strongly in codon usage from the average gene but are relatively similar in codon usage to ribosomal protein genes (RPs) tend to be PHX (8). By these criteria, PHX genes in most prokaryotic genomes include RP genes, translation/transcription processing factors (TF), chaperone/degradation protein complexes (CH), major genes of energy metabolism (EN) (e.g., glycolysis and respiration genes in *Escherichia coli*, photosynthesis genes in *Synechocystis*, methanogenesis genes in *Methanococcus jannaschii*), and essential genes involved in amino acid, nucleotide, and fatty acid biosynthesis.

Codon Usage Contrasts Between Gene Classes

Let G be a family of genes with average codon frequencies $g(x,y,z)$ for the codon nucleotide triplet (x,y,z) normalized so that

$\sum_{(x,y,z)=a} g(x,y,z) = 1$, where the sum extends over all codons (x,y,z) translated to amino acid a . Let $f(x,y,z)$ indicate the average codon frequencies for the gene group F (F can be a single gene g), again normalized to 1 for the codons of each amino acid. The codon usage difference of the gene family F relative to the gene family G (termed the codon bias of F with respect to G) is calculated by the formula:

$$B(F|G) = \sum_a p_a(F) \left[\sum_{(x,y,z)=a} \left| f(x,y,z) - g(x,y,z) \right| \right]$$

where $\{p_a(F)\}$ are the average amino acid frequencies of the genes of F (9, 10). The following gene classes are basic: C = all protein genes; RP = ribosomal protein genes; CH = chaperone and associated degradation genes; and TF = translation/transcription processing factors. Qualitatively, a gene g is PHX (i.e., predicted highly expressed) if $B(g|C)$ is high, whereas $B(g|RP)$, $B(g|CH)$, and $B(g|TF)$ are low. The rationale is reviewed in ref. 8. Predicted expression levels with respect to individual standards are based on the ratios

$$E_{RP}(g) = \frac{B(g|C)}{B(g|RP)}, E_{CH}(g) = \frac{B(g|C)}{B(g|CH)}, E_{TF}(g) = \frac{B(g|C)}{B(g|TF)}.$$

Combined, these produce the expression measure

$$E = E(g) = \frac{B(g|C)}{1/2 B(g|RP) + 1/4 B(g|CH) + 1/4 B(g|TF)}$$

Other weighted combinations can also be used.

Definition I: A gene, g , is predicted to be highly expressed (PHX) if the following two conditions are satisfied: at least two among the three expression values $E_{RP}(g)$, $E_{CH}(g)$, and $E_{TF}(g)$ exceed 1.05 and the overall expression level $E(g)$ is ≥ 1.00 .

Definition II: A gene, g , is predicted to be alien (PA) provided $B(g|RP) \geq M + 0.10$, $B(g|CH) \geq M + 0.10$, $B(g|TF) \geq M + 0.10$, and $B(g|C) \geq M + 0.10$, where M is the median value among $B(g|C)$ for all g .

DEIRA Resistance to Ionizing Radiation and UV Damage

The mesophilic bacterium DEIRA can survive an intense dose of ionizing radiation that is lethal to virtually all other organisms (5, 6). Such ionizing radiation causes DNA strand breaks, generates DNA cross-links, and invokes a myriad of other DNA, RNA, and protein damage. Radiation consequences apply directly to DNA in about 30% of effects and in 70% of effects to

Abbreviations: DEIRA, *Deinococcus radiodurans*; PHX, predicted highly expressed; PA, putative alien; RP, ribosomal protein gene; TF, translation/transcription processing factor; CH, chaperone/degradation proteins; EN, energy metabolism; TCA, tricarboxylic acid.

*To whom reprint requests should be addressed. E-mail: fd.zgg@forysye.stanford.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Selected classes of PHX genes on Chromosome I, designated DR, and Chromosome II (megaplasmid-1), designated DRA

	DR no.	Size, aa	E(g)
Detoxification			
Aconitate hydratase (acn) senses oxidative damage	DR1720	905	2.56
Catalase (katA)	DR1998	535	1.92
sodA	DR1275	210	1.75
Catalase (katE)	DRA0259	771	1.55
Dps (DNA protection during starvation)	DR2263	206	1.51
Tellurium resistance protein terD-1	DR2221	190	1.47
sodC	DR1546	181	1.44
Chloride peroxidase	DR0791	304	1.14
Organic hydroperoxide resistance protein	DR1857	138	1.11
Periplasmic divalent cation tolerance	DR2294	101	1.09
Tellurium resistance terD-2	DR2225	190	1.07
sodC	DRA0202	461	1.05
Chaperones/degradation (export, secretion)			
GroEL	DR0607	547	2.35
DnaK	DR0129	627	2.24
lon1 (ATPdep)	DR1974	812	1.69
PPIase FKBP-type	DR2464	157	1.51
Trigger factor (tig)	DR1948	464	1.48
ftsH-3	DRA0290	655	1.47
PPIase cyclophilin-type	DR0237	192	1.46
lon2	DR0349	820	1.41
HSP20	DR1114	182	1.39
clpX	DR1973	402	1.32
ppiC	DR1063	343	1.27
Phage shock protein A (pap)	DR1473	222	1.26
grpE	DR0128	220	1.24
PPIase cyclophilin-type	DR2542	175	1.23
trxA (thioredoxin)	DR0944	140	1.23
groES	DR0606	120	1.14
msrA	DR1849	205	1.04
Major transcription/translation genes			
EF-Ts	DR1512	263	2.33
EF-G (fus-1)	DR0307	697	2.27
EF-Tu (tuf-2)	DR2050	404	2.23
EF-Tu (tuf-1)	DR0309	404	2.15
rpo-C	DR0911	1545	2.06
GTP binding elongation factor typA/bipA	DR1198	592	1.99
rpoB	DR0912	1179	1.95
infB	DR1799	596	1.72
EF-P	DR0119	184	1.63
Transcription elongation factor (greA-2)	DR1970	160	1.41
greA-1	DR1162	157	1.32
NusG antiterminator	DR2047	189	1.26
N utilization A (nusA)	DR1797	401	1.24
rpoA	DR2128	340	1.19
infC	DR2087	211	1.14
frr (ribosome recycling factor)	DR1510	182	1.12
Cell envelope and surface structure			
S-layer protein	DR2577	1166	2.33
S-layer-like array-related protein	DR0383	279	2.00
S-layer-like array-related protein	DR1185	346	1.79
Immunogenic protein (bcsp31)	DR1649	318	1.61
Hexagonally packed intermediate-layer surface protein (hpi)	DR2508	947	1.58
S-layer-like array-related protein	DR1115	397	1.43
RfbJ protein	DR1080	218	1.27
Putative fimbrial assembly protein PilM	DR0770	390	1.23
Penicillin-binding protein 2	DR1868	436	1.20
Twisting motility protein	DR1963	419	1.18

other cellular components. It was conjectured that, although DEIRA possesses the standard prokaryotic repair repertoire, there must occur intrinsic mechanisms that augment repair (3, 5). It is estimated that at least 40% of damaged DNA is degraded, presumably by controlled cellular exonucleases. Secondly, with the presence of multiple genome equivalents (4 to 10 copies; ref. 11), it is postulated that recombinational processes contribute vitally to reconstituting the cleaved DNA fragments. Under UV radiation, the DEIRA chromosomes generally fracture, but in less than 3 h, the fragments can be reassembled without loss of viability. But, Harsojo *et al.* (11) reported that, for varied genome complements in DEIRA, there is no change in radiation resistance capacities. Actually, *E. coli* and *Vibrio cholerae* during fast growth contain at least three genome equivalents, and it is suggested that the extra chromosomal

equivalents facilitate interchromosomal recombination and repair (3). However, it is enigmatic that the set of repair proteins are predominantly of low to moderate predicted expression levels, excepting RecA and a weak MutT homologue.

We proffer a different explanation for the remarkable UV resistance capacity of DEIRA that is suggested by the nature of the PHX genes of the genome. Four classes of genes stand out (Tables 1, 2, and 3): (i) the abundant degradation and export vehicles available for removing damaged DNA, RNA, and proteins; (ii) the surfeit of chaperones, of which many putatively enhance the operations of the repair proteins; (iii) the manifold detoxification facilities that neutralize and remove free oxygen radicals and other toxic substances; and (iv) the striking abundance of PHX protease genes in DEIRA, more than in any other genome (Table 3). Compared with other complete prokaryotic

Table 2. PHX genes of DEIRA in energy metabolism

	DR no.	Size, aa	E(g)
Glycolysis/gluconeogenesis			
Glyceraldehyde 3-phosphate dehydrogenase (gap)	DR1343	329	1.66
Enolase (eno)	DR2637	421	1.47
Fructose-bisphosphate aldolase (fbaA)	DR1589	304	1.36
Phosphoenolpyruvate carboxykinase (pckA)	DR0977	489	1.29
L-lactate dehydrogenase (ldh)	DR2364	303	1.05
Energy-TCA			
Aconitate hydratase (acn)	DR1720	905	2.56
Isocitrate lyase (aceA)	DR0828	463	2.19
Isocitrate dehydrogenase (icd)	DR1540	430	2.07
Malate dehydrogenase (mdh)	DR0325	329	1.80
Succinyl-CoA synthetase β (sucC)	DR1247	385	1.71
2-oxoglutarate dehydrogenase E1 (sucA)	DR0287	955	1.70
Fumarate hydratase (fumC)	DR2627	812	1.69
Succinyl-CoA synthetase α (sucD)	DR1248	300	1.65
Isocitrate dehydrogenase	DR1674	332	1.56
Citrate synthase (citA)	DR0757	410	1.41
2-oxoglutarate dehydrogenase dehydrolipoamide succinyl transferase E2	DR0083	416	1.31
Succinate dehydrogenase			
(sdhB)	DR0951	263	1.21
(sdhA)	DR0952	582	1.20
(sdhD)	DR0953	124	1.10
Energy: ATP synthase			
v-type ATP synthase-B	DR0701	470	1.90
v-type ATP synthase-K	DR0696	100	1.74
v-type ATP synthase-D	DR0702	223	1.55
v-type ATP synthase-E	DR0697	184	1.47
v-type ATP synthase-I	DR0695	689	1.33
v-type ATP synthase-A	DR0700	581	1.24
Energy: other			
Glycerol-3-phosphate dehydrogenase	DR1019	521	1.80
2-oxo acid dehydrogenase lipoamide dehydrogenase E3	DR2526	480	1.44
Glycerol kinase (glpK)	DR1928	500	1.31

genomes to date, DEIRA contains the greatest number of PHX detoxification genes (for explicit listing, see Table 1). Among 28 complete prokaryotic genomes, numbers of PHX proteases are as follows (in decreasing count): *D. radiodurans*, 14 (see Table 3); *Synechocystis* (PCC6803), 11; *Mycobacterium tuberculosis*, 8; *Pyrococcus abyssi*, 6; *E. coli* (K-12), 5; *Aquifex aeolicus*, 5; *Mycoplasma pneumoniae*, 4; *Ureaplasma urealyticum*, 4; *Thermotoga maritima*, 4; *Bacillus subtilis*, 2; *Rickettsia prowazekii*, 2; *Borrelia burgdorferi*, 2; *Treponema pallidum*, 2; *Thermoplasma acidophilum*, 2; *Pyrococcus horikoshii*, 2; *Pseudomonas aeruginosa*, 1; *Helicobacter pylori* (J99), 1; *Campylobacter jejuni*, 1; *Chlamydia pneumoniae*, 1; *Mycoplasma genitalium*, 1; *Archaeoglobus fulgidus*, 1; and *Halobacterium* (NRC-1), 2. Zero PHX proteases are found in *Chlamydia trachomatis*, *Haemophilus influenzae*, *H. pylori* (26695), *M. jannaschii*, *Methanobacterium thermoautotrophicum*, and *Neisseria meningitidis*.

We propose that the abundance and versatility of PHX chaperone/degradation, protease, and detoxification genes help intrinsically in maintaining the survival and stability of the *D. radiodurans* cell when exposed to severe conditions of UV radiation. Along these lines, there are precedents for chaperone influences that enhance the UvrA function in removing DNA damage by nucleotide excision repair and that facilitate other repair processes as well and have the ability to render them more effective (12). The PHX chaperone, degradation, protease, and detoxification protein ensemble allow DEIRA to maintain the integrity of its essential macromolecules. DEIRA also features three PHX flagellar genes. It is conceivable that these genes play a role in Type III secretion and export of deleterious metabolites.

The predicted expression levels of the main repair proteins of *E. coli* and DEIRA (Table 4) are not high, with one exception. RecA is PHX ($E(g) = 2.04$, dramatically high). A RecA mutant and other recombination-deficient mutants appear to be more sensitive to both UV light and ionizing radiation than wild-type cells (13). RuvB (1.06) and Ssb (1.02) do not precisely qualify as

PHX by Definition I but can be considered marginally PHX. By contrast, the bulk of major chaperone, degradation, and protease proteins are unambiguously PHX.

The DNA repair genes identified in DEIRA all have functional analogs in other prokaryotic species. It is claimed (1) that DEIRA possesses high redundancy in repair, but, apart from the Nudix family of nucleoside triphosphate pyrophosphorylases, this is not confirmed. Actually, only one gene of the “Nudix family” (DR2356) is PHX. Moreover, it is questionable that DR2356 is a homolog of the *E. coli* MutT. The gene is 432 nt (144 aa), and its probability value in a standard BLAST alignment is only 0.021. Mutual similarity among the DEIRA Nudix family proteins is low, mostly less than 15% identity. Moreover, the Nudix family are hardly consistent in sizes and show expression levels in the range $E = 0.3$ to 0.9, and only one gene, DR2356, is PHX with E value 1.29. Paradoxically, two of the Nudix protein family meet the criteria of being putative alien genes (DR0149, DR0784).

Does DEIRA possess novel repair proteins that are more effective than the known proteins in other species? Does it use the common repair machinery more efficiently or in new ways? It is interesting that, among PHX genes in DEIRA, there are 105 ORFs with unknown functions. This number exceeds the percentage of unknown ORFs among PHX genes in most other bacteria (8).

Special PHX Genes of DEIRA

D. radiodurans possesses a collection of PHX S-layer (surface structure) proteins that may provide environmental protection (e.g., against desiccation or thermal effects). There are four PHX S-layer proteins, an essential part of the cell envelope surface structures (see Table 1, “Cell envelope” column). In general, these structures envelop the cell exterior and presumably implement a protective function preventing damage from large

Table 3. PHX protease genes in *D. radiodurans* and *E. coli*,* *E(g)* values†

DEIRA	<i>E. coli</i>	Genes
1.69	(0.91)	ATP-dependent protease LA (lon) (DR1974, DR0349) [§]
1.41	—	Putative carboxyl-terminal protease (DR1551)
1.48	—	Cell division metalloprotease (ftsH) (DRA0290, DR0583, DR1020) [§]
1.47	1.52	
(0.59)	—	
(0.50)	—	Protease I (pfpl) (DR1199)
1.46	—	Serine protease, subtilase family (DR0812)
1.41	—	Serine protease, subtilase family (DRA0283)
1.38	—	Endopeptidase-related protein (DR1749) [§]
1.33	—	
(0.84)	1.04	ATP-dependent Clp protease, ATP-binding subunit (clpX) (DR1973)
1.32	—	Serine protease, subtilase family, C-terminal fragment (DR2322)
1.29	—	Zinc metalloendopeptidase, leishmanolysin family (DRA0249)
1.28	—	Signal peptidase I (DR1427, DR1737, DR1321) ^c
1.13	—	
(0.78)	—	
(0.67)	—	Putative peptidase (DR1346)
1.08	—	Lipoprotein signal peptidase (lspA) (DR2388)
1.02	1.26	Protease DO precursor (htrA) (DR1756)
(0.66)	1.09	Aminoacyl-histidine dipeptidase (pepD)
—	1.32	Putative integral membrane protease (hflC)
—	—	

*Only genes that qualify as PHX in at least one of the two genomes are included.

†Values for genes that do not qualify as PHX are shown in parentheses. Minus signifies the gene does not have a homolog in this genome.

§For genes carrying multiple homologs in the same genome, the *E(g)* values for the homologous genes are given in the next line(s).

hazards in the environment, e.g., from protein proteases excreted by other species or from predators.

Surprisingly (compared with *E. coli*), many ABC transporter genes (e.g., of peptides, branched chain amino acids, phosphates, and various sugars) are PHX. In this context, DEIRA excretes a protease and grows best on rich media (especially digests of complex proteins). The peptide transporters putatively give the cell access to the amino acids generated by this protease.

Intriguingly, the top PHX gene in DEIRA is the multifunctional tricarboxylic acid (TCA) cycle gene aconitase hydratase (aconitase), *E(g)* value = 2.56. Aconitase functions also in sensing oxygen radicals (14, 15). Aconitase apparently also alerts the cell to large amounts of oxidative damage. In the TCA cycle, aconitase interconverts citrate and isocitrate carrying cis-aconitase as an intermediate. Aconitase has an active iron sulfur (4Fe-4S) core at the catalytic site and also regulates gene expression of the iron transport transferrin receptor and iron storage ferritin. This protein is similar to the cytosolic iron regulatory protein (IRP) of eukaryotes. At its iron-sulfur cluster, aconitase can be inactivated by oxidative stress or iron deprivation. In several other species, apo-aconitase (aconitase without the Fe-S cluster) has the ability to bind mRNA, and available evidence indicates that this protein can function as a transcriptional activator (15). On this basis, it has been suggested that aconitase serves as a sensor, detecting changes in the redox state or the levels of iron within the cell. Aconitase is PHX in other genomes: *E. coli* (*E* = 1.40), *B. subtilis* (1.08), *M. tuberculosis* (1.18), and *A. aeolicus* (1.24). FtsH, an integral inner membrane protein, is a cell division metalloprotease that facilitates degradation and protein folding (16). Alkyl hydroperoxide reductase serves to protect the cell from DNA damage caused by alkyl hydroperoxide.

Penicillin-binding proteins are enzymes that establish peptide cross links in the peptidoglycan layer of the cell. Because the peptidoglycan is necessary throughout growth, we could expect penicillin-binding proteins to be constitutively expressed and concomitantly PHX. The penicillin-binding protein 2 (DR 1868) is indeed PHX with *E(g)* = 1.20. Penicillin-binding protein (pbp, pal) of PHX status is found in *B. burgdorferi* (1.13), *C. jejuni* (1.18), *D. radiodurans* (1.20), three in *M. tuberculosis* (1.15, 1.03, 1.30), *R. prowazekii* (1.03–MurA), *Synechocystis* (1.03–MurZ),

and *V. cholerae* (1.80-peptidoglycan-associated lipoprotein precursor, outer membrane protein). DEIRA possesses an inner and outer membrane with a very thick peptidoglycan resembling a Gram-positive bacterium, although, on the basis of 16S rRNA alignments, DEIRA is currently classified as most similar to *Thermus* spp.

In *E. coli*, isocitrate lyase converts isocitrate into succinate and glyoxylate, allowing carbon that entered the TCA cycle to bypass the formation of 2-oxoglutarate and succinyl-CoA. The succinate can reenter the TCA cycle as succinate glyoxylate, with the assistance of malate synthase A. Also, acetyl-CoA reenters the TCA cycle as malate. Isocitrate lyase (aceA) competes with isocitrate dehydrogenase for isocitrate, and the allosteric regulation of both proteins is well documented. When extant, these proteins play an important role in carbon metabolism of *E. coli*. Among the current complete genomes, isocitrate lyase is PHX only in DEIRA, *M. tuberculosis*, and *P. aeruginosa*; is present, but not PHX, in *E. coli* and *V. cholerae*; and is missing from the other complete genomes (8).

The TCA cycle isocitrate dehydrogenase (icd) gene is PHX in *E. coli* (*E(g)* = 1.65), *B. subtilis* (1.33), *D. radiodurans* (two copies 2.07, 1.56), *H. pylori* (1.11), *M. tuberculosis* (two copies 1.25, 0.92), *A. aeolicus* (1.13), *T. maritima* (1.15), *A. fulgidus* (1.05), marginally PHX in *R. prowazekii* (0.95) and *M. thermoautotrophicum* (1.01), but not PHX in *C. jejuni* (0.76), and *M. jannaschii* (0.85, 0.68). The isocitrate dehydrogenase gene was not found in *H. influenzae*, *C. trachomatis*, *M. genitalium*, *M. pneumoniae*, *B. burgdorferi*, *T. pallidum*, *P. horikoshii*, and *Pyrococcus aerophilum*.

The TCA cycle not only produces ATP energy but significantly feeds into making precursors, macromolecules, and numerous carbon intermediates. Predicted expression levels of the DEIRA TCA cycle genes are similar to *E. coli* other than the exceptions mentioned earlier.

Cyclophilin proteins are remarkably conserved through evolution. Moreover they have been found in almost all organisms, which attests to their fundamental importance (17–19). Because of their enzymatic properties, multiplicity, cellular localization, and role in protein folding, they behave as molecular chaperones. All of the proteins of the cyclophilin family possess enzymatic peptidyl-prolyl cis-trans isomerase activity (PPI-ase), which ac-

Table 4. Repair proteins of *E. coli* and *D. radiodurans*

<i>E(g)</i>		
<i>E. coli</i>	<i>D. radiodurans</i>	Gene
Direct Repair:		
0.38	—	Photorepair protein (PhrA)
0.40	—	Deoxyribodipyrimidine photolyase (PhrB)
0.51	—	O6-methylguanine-DNA-alkyltransferase (Ogt)
0.53	—	Methyltransferase, alkyltransferase (Ada)
Base excision repair:		
0.47	0.56	Uracil DNA glycosylase (Ung)
0.53	0.65	A/G-specific adenine glycosylase (MutY)
0.59	0.64	G/U mismatch-specific DNA glycosylase (mug)
0.57	0.86, 0.79	Endonuclease III (Nth)
0.53	0.57	Formamidopyrimidine-DNA glycosylase (MutM)
0.55	—	Endonuclease VIII (Nei)
0.61	—	3-methyladenine DNA glycosylase I (Tag)
0.56	—	DNA-3-methyladenine glycosidase II (AlkA)
—	0.52	UV damage endonuclease, putative (DR1819)
AP endonucleases:		
0.64	0.66	Exonuclease III (XthA)
0.53	—	Endonuclease IV (Nfo)
Mismatch excision repair:		
0.46	0.45	DNA mismatch repair (MutS)
0.40	0.54	DNA mismatch repair (MutL)
0.63	—	Patch repair protein (Vsr)
0.44	—	GATC endonuclease (Muth)
0.43	—	GATC methylase (Dam)
0.46	—	Exodeoxyribonuclease large subunit (XseA)
0.85	—	Exodeoxyribonuclease small subunit (XseB)
Nucleotide Excision repair:		
0.66	0.76, 0.52	Excision nuclease (UvrA)
0.57	0.55	Excision nuclease (UvrB)
0.40	0.57	Excision nuclease (UvrC)
0.68	0.52	DNA helicase II (UvrD)
0.38	0.58	Transcription-repair coupling factor (Mfd)
Recombinational repair:		
0.28	—	exoV helicase (RecB)
0.32	—	exoV nuclease (RecC)
0.39	0.49	exoV helicase (RecD)
0.51	0.58	Single-stranded DNA binding protein (RecF)
0.39	0.50	5'→3' single stranded DNA exonuclease (RecJ)
0.59	—	Stimulates binding of RecF (RecO)
0.71	0.80	Stimulates binding of RecF (RecR)
0.56	0.51	ATP binding (RecN)
0.45	0.40	DNA helicase (RecQ)
0.73	—	Exodeoxyribonuclease VIII (RecE)
0.68	—	Binds ssDNA and promotes renaturation (RecT)
0.40	—	Exodeoxyribonuclease I (SbcB)
0.35	0.71	dsDNA exonuclease (SbcC)
0.45	0.55	dsDNA exonuclease (SbcD)
0.50	—	DNA repair protein (RadC)
*1.48	*2.04	Recombinase (RecA)
Branch migration/resolution:		
0.59	0.69	Binds specifically to Holliday junctions (RuvA)
0.56	1.06	ATPase, promotes branch migration (RuvB)
0.48	0.50	ATPase, binds Holliday junctions (RecG)
0.61	0.88	Holliday junctions endonuclease (RuvC)
DNA ligases:		
0.56	0.61	NAD-dependent DNA ligase (DnIJ)
Direct Repair:		
Nucleotide pools:		
0.72	0.80	GTPase triphosphatase (MutT)
0.97	—	Nucleotidohydrolase (Dut)
Replication:		
0.48	0.57	DNA polymerase I (PolA)
0.43	—	DNA polymerase II (PolB)
0.64	0.69	DNA polymerase III, alpha chain (DnaC)
Other repair proteins:		
0.47	—	SOS mutagenesis (UmuC)
0.64	—	SOS mutagenesis (UmuD)
0.50	0.87	DNA repair protein (RadA)
0.40	—	Primosomal protein replication factor (PriA)
*1.48	1.02	Single-strand DNA-binding protein (Ssb)
0.99	0.52	ATP-dependent helicase (HepA)
0.54	0.83	Site-specific integrase/recombinase (XerD)
0.54	—	Integrase/recombinase (XerC)
—	0.61	Extracellular nuclease, putative (DRB0067)
—	0.65	Putative competence-damage protein (CinA)

**E(g)* values labeled by an asterisk qualify as PHX by Definition I.

celerates protein folding *in vivo*. Recently PPI-ase activity was suggested as playing a role in regulation of transcription and differentiation. However, not all cyclophilin functions are explained by PPI-ase activity. For instance, one of the cyclophilins plays a regulatory role in the heat shock response. Also, the mitochondrial cyclophilin (Cyclophilin D), an integral part of the mitochondrial permeability transition complex, is regarded as having a crucial role in mechanisms of cell death. In support of a role in the stress response, cyclophilins have recently been shown to be up-regulated under various stressful conditions (18, 19). Two cyclophilin PPIases are PHX in DEIRA. We propose that the cyclophilin proteins and other chaperones contribute importantly to the DEIRA genome capacities for resistance to ionizing and radiation damage.

The "giant" ribosomal protein S1 of DEIRA is critical to all protein synthesis within the cell. This is the second largest ribosomal protein (628 aa) among all current complete prokaryotic genomes. (The S1 protein of *Treponema pallidum* has a length of 820 aa.)

Protein abundances of the two genes, katE and tuf, assessed by mass spectroscopy, record very high levels (M. Lipton, Pacific Northwest National Laboratory, Richmond, WA; personal communication). These genes carry the consistent strong PHX values 1.92 and 2.23, respectively.

Genes not PHX in DEIRA include many of transcriptional regulation, of amino acid and nucleotide biosynthesis, the latter partly because DEIRA engages in scavenger activity.

PA Genes of DEIRA

Among the PA genes of DEIRA, a cluster of enzymes contributing to lipopolysaccharide biosynthesis stands out. These are surface antigens that help to protect the bacterium. Lipopolysaccharide biosynthesis genes are frequently PA in other bacterial genomes (20). Lipopolysaccharide biosynthesis proteins are also replete with repeat structures that may help confound host immune reactions.

There are six Par proteins (which play a role in cell partitioning) in the PA list (see <ftp://gnomic.stanford.edu/pub>) and three type IV pilin genes (DR0548, DR1232, and DR1233). In DEIRA, we can identify 41 individual transposases (nine families), of which 28 qualify as PA genes. Most PA genes are of unassigned function.

The collection of transposases, each of length exceeding 80 aa in DEIRA, numbers 25 on Chromosome I (of which 16 are PA), 1 on Chromosome II (0 PA), 10 encoded on the megaplasmid-2 (8 PA), and 5 relegated to the small plasmid (4 PA). The bulk among transposases have length 407 aa or 415 aa. A smaller family have length 326 aa. There are three Resolvase genes (1 in

Chromosome I, 1 in Chromosome II), both PA, contributing in transposon related function. It is recognized that transposable elements can provide genomic flexibility and dynamic exchange.

The total number of PA genes in DEIRA is 282 (9% of all genes), of which 190 are ORFs of unknown function.

Clusters of PHX and PA Genes

Major PHX gene clusters include an operon of ribosomal proteins, with the EF-G (fus-1) and EF-Tu (tuf-1) genes interspersed extending about 12 kb on Chromosome I from about position 307 kb to position 320 kb. A second PHX cluster consists of the ν -type ATP synthase genes, subunits I, K, E, A, B, and D in the region 710–719 kb. A third PHX cluster involves several TCA and electron transport genes covering 959–967 kb. The subsequent cluster contains four contiguous peptide ABC transporters 1585–1592 kb; then two additional ribosomal protein operons 2058–2063 kb (L7/L12, L10, L1, and L11, followed by genes NusG and EF-Tu (tuf-2)) and another RP operon 2119–2133 kb (S8, L6, L18, S5, L15, S13, S11, S4, and L17, the latter cluster including rpoA, the DNA-directed RNA polymerase alpha subunit). There is a PHX cluster of three ORFs DR2179 [$E(g) = 1.44$], DR2180 [$E(g) = 1.54$], and DR2181 [$E(g) = 1.08$] proximal to position 2178 kb. A PA cluster on Chromosome II features seven genes of lipopolysaccharide biosynthesis.

Concluding Comments

DEIRA features a strikingly high representation of PHX chaperone, detoxification, protease genes, and several proteins of cell envelope surface structures. We propose that many of these contribute significantly in radiation resistance and cellular remediation. For example, DEIRA chaperone/degradation and protease proteins play an enhance role in degradation and export of damaged DNA and proteins. Also DEIRA chaperones facilitate resistance to radiation by altering the operating conditions of many enzymes. The multiple PHX detoxification genes of Catalase and SOD help remove active radicals that may be generated by radiation. Many cell division proteins are PHX, which could speed up cell growth. Also, manganese possibly entering via transporters can induce rapid cell division in *Deinococcus* culture (21).

The number of PHX proteins of unknown function (105) is strikingly high. Many of these might be novel proteins contributing vitally in radiation resistance.

We are happy to thank Drs. D. Bhaya, G. Miklos, and J. Trent for constructive comments on the manuscript. This work was supported in part by National Institutes of Health Grants 5R01GM10452-35 and 5R01HG00335-13.

- White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L., *et al.* (1999) *Science* **286**, 1571–1577.
- Udupa, K. S., O'Cain, P. A., Mattimore, V. & Battista, J. R. (1994) *J. Bacteriol.* **176**, 7439–7446.
- Minton, K. W. & Daly, M. J. (1995) *BioEssays* **17**, 457–464.
- Daly, M. J. & Minton, K. W. (1995) *J. Bacteriol.* **177**, 5495–5505.
- Battista, J. R. (1997) *Annu. Rev. Microbiol.* **51**, 203–224.
- Battista, J. R., Earl, A. M. & Park, M. J. (1999) *Trends Microbiol.* **7**, 362–365.
- Minton, K. W. (1996) *Mutat. Res.* **363**, 1–7.
- Karlin, S. & Mrázek, J. (2000) *J. Bacteriol.* **182**, 5238–5250.
- Karlin, S., Mrázek, J. & Campbell, A. M. (1998) *Mol. Microbiol.* **29**, 1341–1355.
- Karlin, S., Campbell, A. M. & Mrázek, J. (1998) *Annu. Rev. Genet.* **32**, 185–225.
- Harsojo, S., Kitayama, S. & Matsuyama, A. (1981) *J. Biochem. (Tokyo)* **90**, 877–880.
- Zou, Y., Crowley, D. J. & Van Houten, B. (1998) *J. Biol. Chem.* **273**, 12887–12892.
- Kitayama, S., Narumi, I., Kikuchi, M. & Watanabe, H. (2000) *Mutat. Res.* **461**, 179–187.
- Hentze, M. W. & Kuhn, L., C. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8175–8182.
- Rouault, T. A. & Klausner, R. D. (1996) *Trends Biochem. Sci.* **21**, 174–177.
- Pfanner, N. (1999) *Curr. Biol.* **9**, R720–R724.
- Kunz, J. & Hall, M. N. (1993) *Trends Biochem. Sci.* **18**, 334–338.
- Andreeva, L., Heads, R. & Green, C. J. (1999) *Int. J. Exp. Pathol.* **80**, 305–315.
- Doyle, V., Virji, S. & Crompton, M. (1999) *Biochem. J.* **341**, 127–132.
- Lawrence, J. G. & Ochman, H. (1997) *J. Mol. Evol.* **44**, 383–397.
- Zhang, Y. M., Wong, T. Y., Chen, L. Y., Lin, C. S. & Liu, J. K. (2000) *Appl. Environ. Microbiol.* **66**, 105–112.